

Genetic Incorporation of *N*^ε-Formyllysine, a New Histone Post-translational Modification

Tianyuan Wang,^[a] Qing Zhou,^{*,[b]} Fahui Li,^[b] Yang Yu,^[c] Xuebin Yin,^{*,[a, d]} and Jiangyun Wang^{*,[b]}

Lysine formylation is a newly discovered post-translational modification (PTM) in histones and other nuclear proteins; it has a well-recognized but poorly defined role in chromatin conformation modulation and gene expression. To date, there is no general method to site-specifically incorporate *N*^ε-formyllysine at a defined site of a protein. Here we report the highly efficient genetic incorporation of the unnatural amino acid *N*^ε-formyllysine into proteins produced in *Escherichia coli* and mammalian cells, by using an orthogonal *N*^ε-formyllysine tRNA-synthetase/tRNA_{CUA} pair. This technique can be applied to study the role of lysine formylation in epigenetic regulation.

Post-translational modification (PTM) of histones is vital for the regulation of diverse biological processes, including DNA replication, DNA repair, and maintenance of genomic stability.^[1] Aberrant histone modification leads to many diseases in human.^[2] To reveal the role of a specific PTM in histones, it is essential to produce homogeneous recombinant protein that contains the site-specific PTM. Semisynthetic and enzyme-catalyzed methods have been developed to realize site-specific histone PTM.^[3] Compared to these methods, the genetic code expansion technique has the unique advantage that unnatural amino acids (UAAs) can be directly incorporated at specific sites of a protein of interest.^[4] For example, three lysine PTMs (*N*^ε-acetyllysine (AcK), *N*^ε-crotonyl-

lysine, and *N*^ε-methyllysine) have been successfully genetically encoded.^[5] Lysine formylation is a newly discovered PTM in histones and other nuclear proteins, and it is believed to be associated with oxidative stress under pathological conditions.^[6] The formyl moiety can come from 3'-formylphosphate residues arising from 5'-oxidation of deoxyribose in DNA, caused by the enediyne neocarzinostatin. Only one methyl group shorter than AcK,^[7] *N*^ε-formyllysine (ForK) is not only structurally similar, but also appears at the same sites of histones. This raises the question as to whether ForK and AcK have similar roles in chromatin structure modulation and gene expression. However, the lack of methods to site-specifically incorporate ForK into defined sites of proteins limits our ability to probe the function

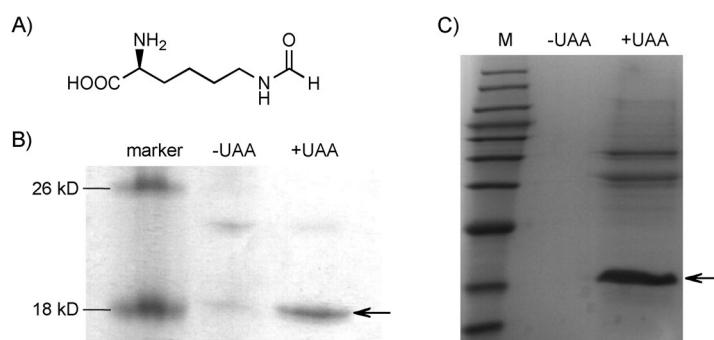


Figure 1. Site-specific ForK incorporation. A) Chemical structure of ForK. B) SDS-PAGE analysis of myoglobin-TAG4-ForK expression in the presence (+UAA) or absence (–UAA) of 1 mM ForK. C) SDS-PAGE analysis of the expression of histone H3-TAG23 ForK in the presence (+UAA) or absence (–UAA) of 1 mM ForK.

[a] T. Wang, Prof. X. Yin

School of Earth and Space Science
University of Science and Technology of China (USTC)
96 Jinzhai Road, Hefei, Anhui 230026 (China)
E-mail: xbyin@ustc.edu.cn

[b] Dr. Q. Zhou, Dr. F. Li, Prof. J. Wang

Laboratory of RNA Biology
Institute of Biophysics, Chinese Academy of Sciences
15 Datun Road, Chaoyang District, Beijing 100101 (China)
E-mail: zqboy008@163.com
jwang@ibp.ac.cn

[c] Dr. Y. Yu

Institute of Industrial Biotechnology, Chinese Academy of Sciences
32 West 7th Avenue, Tianjin Airport Economic Area, Tianjin 300308 (China)

[d] Prof. X. Yin

Suzhou Key Lab for Selenium and Human Health
Suzhou Institute for Advanced Study, USTC
166 Ren'ai Road, Dushu Lake Higher Education Area, Industrial Park
Suzhou, Jiangsu 215123 (China)



Supporting information for this article is available on the WWW under
<http://dx.doi.org/10.1002/cbic.2.201500170>.

of ForK in vitro and in vivo.

Here we report the highly efficient genetic incorporation of the UAA ForK (Figure 1 A) into *Escherichia coli* and mammalian cells by using an orthogonal *N*^ε-formyllysine tRNA synthetase/tRNA_{CUA}^{Pyl} pair. We found that ForK-bearing proteins are not recognized by anti-AcK antibody, thus indicating that the in vivo proteins bearing ForK might interact with other proteins differently from those bearing AcK.

To selectively incorporate ForK at defined sites in proteins, we performed three rounds of positive and two rounds of negative selection with a *Methanosarcina barkeri* pyrrolysyl-tRNA-synthetase (*MbPylRS*) library, in order to evolve an orthogonal tRNA/aminoacyl-tRNA synthetase pair that selectively charges ForK in response to the amber (TAG) codon, as previously described.^[4i, 8] The selected ForK-specific PylRS ("ForKRS") had five mutations: L266M, L270I, Y271F, L274A, and C313F. ForKRS was digested and two copies of ForKRS gene were ligated sequen-

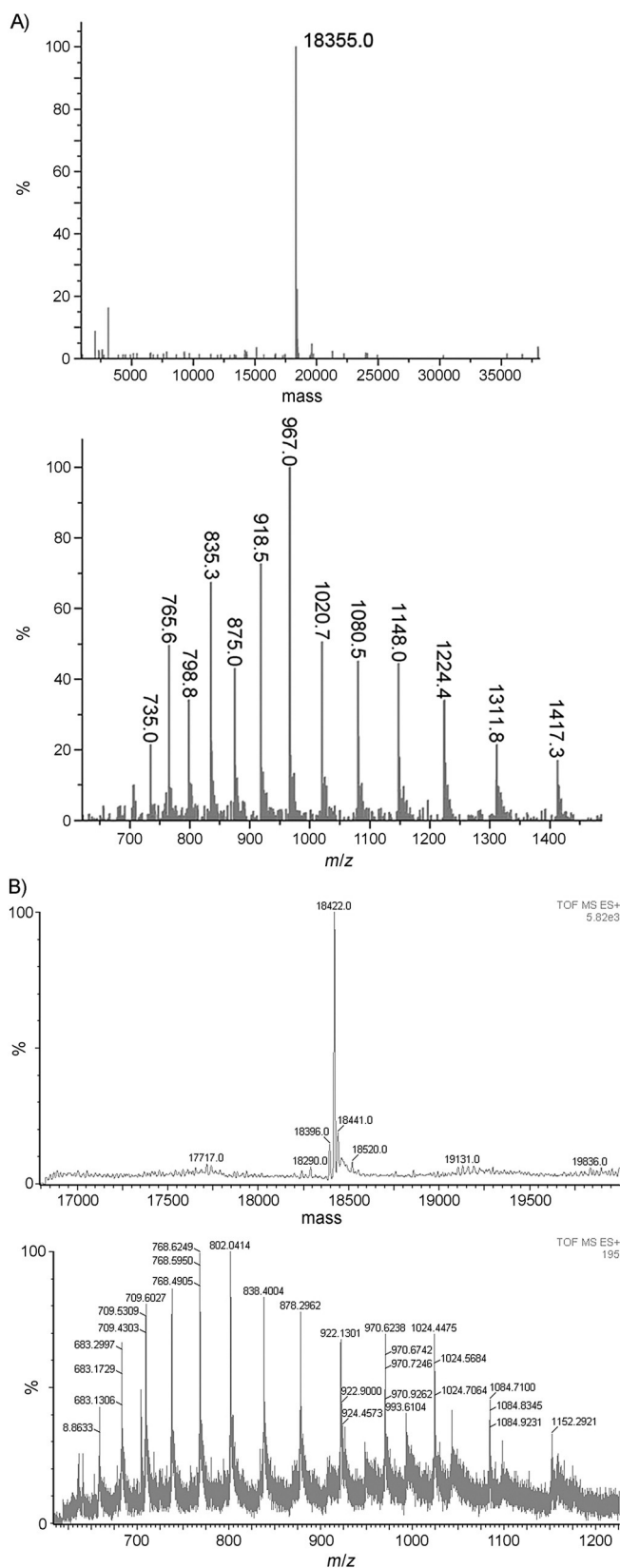


Figure 2. ESI-MS analysis of myoglobin. A) Wild-type (calcd: 18354 Da; found: 18355.0 Da). B) Myoglobin-TAG4-ForK (calcd: 18423.0 Da; found: 18422.0 Da).

tially into Sall/BglII and Pst1/Nde1 restriction sites of the pEVOL vector^[9] to generate pEVOL1. Then, *Methanosarcina barkeri* pyrrolysyl-tRNA_{CUA}(pylT) was ligated into the Apa1/Xho1 sites of pEVOL1 to generate pEVOL-ForkRS, which was used to test the efficiency and selectivity of ForkRS, as previously reported.^[4c]

To determine whether Fork can be incorporated into a protein with high efficiency and fidelity, an amber stop codon was substituted for Ser4 in sperm whale myoglobin to construct the myoglobin-TAG4 expression vector. Protein expression was carried out in *E. coli* in the presence of the selected synthetase (ForkRS), *MbtRNA*^{Pyl}_{CUA}(pylT) and 1 mM Fork (or in the absence of Fork as a negative control). Analysis of the purified protein by SDS-PAGE showed that full-length myoglobin was expressed only in the presence of Fork (Figure 1B), thus indicating that ForkRS recognizes Fork specifically. ESI-MS analysis of myoglobin-TAG4-Fork gave an observed average mass of 18422.0 Da (Figure 2), which was in agreement with the calculated mass (18423 Da). This indicates that Fork had been incorporated at the defined site of myoglobin. Fork was also incorporated into the position 23 of histone H3 (Figure 1C); this has been identified as a lysine formylation and acetylation site.^[6a] The yield of "H3-TAG23-Fork" was 2 mg L⁻¹, which is comparable to yields reported for the incorporation of other UAAs with the pylRS/*MbtRNA*^{Pyl}_{CUA} pair.^[5]

To examine whether Fork is recognized by an anti-AcK antibody, an amber stop codon was substituted for Lys3 and Tyr151 in superfolder green fluorescent protein (sfGFP) to construct the sfGFP-TAG3 and sfGFP-TAG151 expression vectors, respectively. AcK or Fork was incorporated into position 151 of sfGFP, and analyzed by western blotting. sfGFP-TAG3-AcK was detected efficiently by an anti-AcK antibody; sfGFP-TAG3-Fork was not (Figure 3). This implies that lysine formylation PTM is

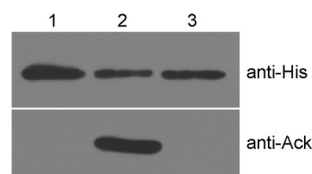


Figure 3. Western blot analysis of wild-type and mutant sfGFP: lane 1) wild-type sfGFP; lane 2) sfGFP-TAG3-AcK; lane 3) sfGFP-TAG3-Fork. The samples were probed by using anti-his tag and anti-N^ε-acetyllysine antibodies.

a significantly different to lysine acetylation.

As the pyrrolysine tRNA synthetase/tRNA pair is orthogonal in both bacterial and mammalian cells,^[10] we tested whether the ForkRS and *MbtRNA*^{Pyl}_{CUA} pair could perform site-specific incorporation of Fork in mammalian cells. We cloned ForkRS into the pCMV-NBK-1 vector^[10] to generate pCMV-ForkRS, in which transcription of *MbtRNA*^{Pyl}_{CUA} is under the control of the human U6 promoter and the expression of ForkRS is driven by the CMV promoter. Plasmids pSwan-EGFP37TAG^[4b] and pCMV-ForkRS were co-transfected into 293T human embryonic kidney cells. The cells were grown in the absence or presence of 1 mM Fork for 36 h. EGFP fluorescence was observed only

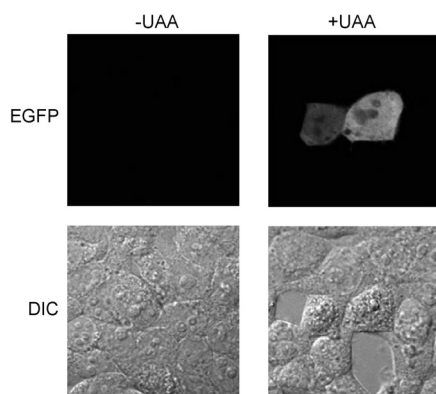


Figure 4. Genetic incorporation of ForK at position 37 of EGFP in 293T cells by using the ForKRS/*MbtRNA*^{Pyl}_{CUA} pair. The EGFP fluorescence is observed only in the presence of 1 mM ForK.

for cells grown in the presence of ForK (Figure 4), thus indicating that when using the ForKRS/*MbtRNA*^{Pyl}_{CUA} pair, ForK was site-specifically introduced at position 37 of EGFP through amber codon suppression in mammalian cells.

Conclusion

We have demonstrated the highly efficient genetic incorporation of ForK into *E. coli* and human cells. The ForK-incorporated protein was not recognized by anti-AcK antibody, thus indicating that proteins bearing PTM lysine formylation might interact with other proteins differently from those bearing lysine acetylation.^[1a,11] Lysine formylation was also demonstrated in histone at the sites for lysine acetylation and methylation. This indicates that lysine formylation can block lysine acetylation and methylation in vivo. Lysine formylation can also negatively regulate nucleosome assembly.^[6a] In our work, ForK was site-specifically incorporated into histones for the first time. This new technology could be powerful tool to elucidate the cellular functions of lysine formylation, caused by oxidative stress in cells. An anti-ForK antibody is not commercially available. By expressing homogeneous recombinant proteins bearing ForK at a specific position, we are currently working towards producing antibodies for the recognition of ForK.

Acknowledgements

We gratefully acknowledge the Major State Basic Research Program of China (2015CB856203), the CAS grant (KJZD-EW-L01), and the National Science Foundation of China (91440116, 21325211, 91313301, 31370016, 21102172). The authors declare that they have no conflict of interest. This article does not con-

tain any studies with human or animal subjects performed by any of the authors.

Keywords: amino acids • formyllysine • gene technology • histones • post-translational modifications

- [1] a) B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41–45; b) X.-J. Yang, E. Seto, *Mol. Cell* **2008**, *31*, 449–461; c) A. L. Burlingame, X. Zhang, R. J. Chalkley, *Methods* **2005**, *36*, 383–394; d) F. Chu, D. A. Nusinow, R. J. Chalkley, K. Plath, B. Panning, A. L. Burlingame, *Mol. Cell. Proteomics* **2006**, *5*, 194–203.
- [2] a) G. Egger, G. Liang, A. Aparicio, P. A. Jones, *Nature* **2004**, *429*, 457–463; b) A. Portela, M. Esteller, *Nat. Biotechnol.* **2010**, *28*, 1057–1068.
- [3] M. Shogren-Knaak, H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson, *Science* **2006**, *311*, 844–847.
- [4] a) J. W. Chin, *Annu. Rev. Biochem.* **2014**, *83*, 379–408; b) X. Liu, J. Li, J. Dong, C. Hu, W. Gong, J. Wang, *Angew. Chem. Int. Ed.* **2012**, *51*, 10261–10265; *Angew. Chem.* **2012**, *124*, 10407–10411; c) F. Li, H. Zhang, Y. Sun, Y. Pan, J. Zhou, J. Wang, *Angew. Chem. Int. Ed.* **2013**, *52*, 9700–9704; *Angew. Chem.* **2013**, *125*, 9882–9886; d) C. Hu, S. I. Chan, E. B. Sawyer, Y. Yu, J. Wang, *Chem. Soc. Rev.* **2014**, *43*, 6498–6510; e) M. J. Schmidt, D. Summerer, *Angew. Chem. Int. Ed.* **2013**, *52*, 4690–4693; *Angew. Chem.* **2013**, *125*, 4788–4791; f) M. J. Schmidt, J. Borbas, M. Drescher, D. Summerer, *J. Am. Chem. Soc.* **2014**, *136*, 1238–1241; g) E. Arbely, J. Torres-Kolbus, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2012**, *134*, 11912–11915; h) D. P. Nguyen, M. Mahesh, S. J. Elsässer, S. M. Hancock, C. Uttamapinant, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 2240–2243; i) Z. Yu, Y. Pan, Z. Wang, J. Wang, Q. Lin, *Angew. Chem. Int. Ed.* **2012**, *51*, 10600–10604; *Angew. Chem.* **2012**, *124*, 10752–10756.
- [5] a) C. H. Kim, M. Kang, H. J. Kim, A. Chatterjee, P. G. Schultz, *Angew. Chem. Int. Ed.* **2012**, *51*, 7246–7249; *Angew. Chem.* **2012**, *124*, 7358–7361; b) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234; c) H. Neumann, S. M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes, J. W. Chin, *Mol. Cell* **2009**, *36*, 153–163; d) H.-w. Ai, J. W. Lee, P. G. Schultz, *Chem. Commun.* **2010**, *46*, 5506–5508; e) Z. U. Wang, Y.-S. Wang, P.-J. Pai, W. K. Russell, D. H. Russell, W. R. Liu, *Biochemistry* **2012**, *51*, 5232–5234.
- [6] a) J. R. Wiśniewski, A. Zougman, M. Mann, *Nucleic Acids Res.* **2008**, *36*, 570–577; b) T. Jiang, X. Zhou, K. Taghizadeh, M. Dong, P. C. Dedon, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 60–65; c) B. Edrissi, K. Taghizadeh, P. C. Dedon, *Plos Genet.* **2013**, *9*, e1003328.
- [7] J. Zhang, R. Sprung, J. Pei, X. Tan, S. Kim, H. Zhu, C.-F. Liu, N. V. Grishin, Y. Zhao, *Mol. Cell. Proteomics* **2009**, *8*, 215–225.
- [8] Q. Zhou, M. Hu, W. Zhang, L. Jiang, S. Perrett, J. Zhou, J. Wang, *Angew. Chem. Int. Ed.* **2013**, *52*, 1203–1207; *Angew. Chem.* **2013**, *125*, 1241–1245.
- [9] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361–374.
- [10] P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem. Int. Ed.* **2009**, *48*, 4052–4055; *Angew. Chem.* **2009**, *121*, 4112–4115.
- [11] T. Kouzarides, *Cell* **2007**, *128*, 693–705.

Manuscript received: March 31, 2015

Accepted article published: April 24, 2015

Final article published: May 7, 2015